

# **Finding New Ways to Read Digital Talking Books**

Miguel Rodrigues  
Luís Carriço  
Carlos Duarte

DI-FCUL

TR-2005-13

July 2005

Departamento de Informática  
Faculdade de Ciências da Universidade de Lisboa  
Campo Grande, 1749-016 Lisboa  
Portugal

Technical reports are available at <http://www.di.fc.ul.pt/tech-reports>. The files are stored in PDF, with the report number as filename. Alternatively, reports are available by post from the above address.



# Finding New Ways to Read Digital Talking Books

Miguel Rodrigues, Luís Carriço, Carlos Duarte

Departamento de Informática  
Faculdade de Ciências da Universidade de Lisboa  
Campo Grande, 1749-016 Lisboa, Portugal

July 2005

## Abstract

In this paper we present new ways to read Digital Talking Books (DTBs). DTBs aim to provide better access to literary contents for the visually impaired population. However, these books offer limited capabilities when reading the book freely. Only continuous reading or jumping through sections of the text is allowed. This lack of flexibility leaves out many of the types of reading users usually do, such as those that provide the user with an overview of the book content or readings based on a specific set of key points that have more importance to the user. The work we present here overcomes these limitations by providing the user with a whole new range of interaction possibilities with DTBs.

## 1 Introduction

With the ever increasing need to make written knowledge available to every person, new efforts have come to light bringing together the new technologies and the traditional ways of content presentation. Digital Talking Books (DTBs) [1] aim to provide access to this knowledge for a print-disabled or visually impaired community, using visual reproduction of the book together with spoken reproduction. This combination of different media manages to overcome some of the cognitive limitations of the human perception [5], helping people obtain a higher range of interaction possibilities.

The possibilities made available by the digital nature of the books, with text and audio accessible simultaneously, not only allow overcoming the limitations imposed by analogue audio books, but also enable exploration and interaction in different, and possibly improved, ways than those allowed by printed material. Improvements like the increased ease of following a cross-reference are complemented by new possibilities, such as automated word searching. Marking, annotating and categorizing parts of a digital book will, at least, streamline the user effort, and will also make easier to share user's comments, and perhaps, even create new reading patterns. Still, managing all these new interaction and navigation options with a mainly audio oriented basis can be a problem, requiring additional ways to interact with the user in audio mode only form. The combination of visual and audio components and the weight each has in

the interaction greatly depends on each person using the player. As such, the possibility to personalize the interface by changing font size, colour, disposition of elements on screen or even changing the audio reproduction speed becomes increasingly important. This last feature assumes crucial importance specially when reading technical or reference books. Most of these aren't usually read in a continuous way and some parts are read several times over. For this type of reading, special features such as different reading speeds, jumping unimportant sections or marking portions of the text for future reference, assume bigger relevance and deserve an attention that they haven't yet received from current DTB players.

This paper presents RiPlay (Rich Digital Talking Book Player). It is a DTB player that fulfils all the requirements for DTB reading, emphasizing the usability and accessibility issues. The interaction and presentation alternatives it offers enable its usage in pure audio mode (speech reproduction and interaction), still maintaining the navigation and annotation capabilities and its ease of use. Apart from a full DTB player, RiPlay offers a couple of mechanisms that target the studying of documental material, namely technical and scientific articles and books. Fast overview reading and listening is offered, through a set of simple text selection criteria, highly customizable. It is also possible to mark text excerpts, with different importance degrees or even small annotations, that later, on a document revision can be read (or listened to) selectively, using different reading speeds or skipping uninteresting parts. In this sense, RiPlay provides the ability to listen to different perspectives of a text.

The paper begins with an overview of related work where the requirements are also raised. Then the project where RiPlay was developed is briefly presented. Section 4 describes the player emphasizing its use in studying activities. Finally conclusion and future work are drawn.

## 2 Related Work

Digital Talking Books (DTBs) are the digital counterpart of analogue talking books, capable of overcoming several of the analogue media limitations. Books on analogue support are linear presentations, which exhibit great shortcomings when reading materials are accessed randomly, like reference works or magazines. Digital media, on the other hand, offers readers the possibility to move around in a book as freely as in a printed book, opening up new interaction possibilities like placing bookmarks or highlighting material. The DTB goes beyond the limits imposed on analogue books, and printed books, including the audio recording of the work, the full text content and images as well.

DTB developments over the last years lead to the appearance of several different specifications, with the Daisy Consortium (<http://www.daisy.org/>) being responsible for the publication of several standards. Finally, in 2002, with co-operation from the Daisy Consortium, the National Information Standards Organization (NISO) published the current standard ANSI/NISO z39.86 [1]. The standard focuses on the structure and content of the files needed to produce DTBs. An auxiliary document, the Playback Device Feature List [9], created during the development of the standard, describes the main features that playback devices should possess.

The essential set of features defined on the Playback Device Feature List

document [9] include:

- no need to use visual display to operate device;
- variable playback speed;
- document accessible at fine level of detail;
- usable table of contents;
- easy skips (moving sequentially through the elements);
- ability to move directly to a specific target;
- reading of notes;
- setting and labelling bookmarks;
- automatic bookmark at stop;
- ability to add information (highlighting and notes);
- fast forward and fast reverse;
- availability of human and electronic speech;
- presentation of visual elements in alternative formats (speech).

The list is rather extensive, even though the document is not normative and does not present specific implementation solutions.

The Daisy Consortium publishes and maintains a list of DTB players in <http://www.daisy.org/tools/playback.asp>. The players adopt different solutions for the presentation and interaction. According to a heuristic evaluation presented in [3], most of them adhere to the navigation related recommendations, with some support for bookmarks and annotations.

Nevertheless, several usability and accessibility flaws are reported that hinder the reading experience: lack of mechanisms to make the reader aware of the presence of annotations in a given part of the book; narration's interruption caused by the examination of auxiliary structures, like the table of contents; font sizes that render the text in such a way that it exceeds the available space in a non-scrollable window; impossibility to customize the presentation (size and colour) of some elements; and an overall impossibility to use the players in a non-visual environment for all but the simpler operations, due mainly to the absence of speech recognition and the consequent over-reliance on keyboard shortcuts.

The possibilities opened by annotating and marking the books are left unexplored by all the players. Assigning degrees of importance to each annotation could be used to control playback speeds. Even if almost all players allow for different playback speeds, none uses information entered by the reader to enable this kind of behaviour. If the user wishes to listen to some segment of the text with greater attention, she may wish to switch the playback to a slower speed. In the current players, that has to be done directly by the user, breaking the readers' concentration. Another issue raised in [3] relates to user awareness of a bookmark or an annotation. Several possibilities to alert the user to the

presence of a bookmark or annotation can be envisaged, from text highlight to the use of other modalities (like sounds and speech).

Furthermore, if reading of technical or reference material is considered, even more limitations become visible. This kind of material is usually not read in a linear fashion, with the user typically skimming over the segments considered less important. Fast and selective reading [4, 8] are important activities, that allow users to decide about the text interest before engaging in a deeper study. None of the players provide any support for these reading patterns.

### 3 The Rich DTB Projects

The work described in this paper was developed within a couple of R&D projects aiming at the construction and usage of rich DTBs. The first one, IPSOM, aimed primarily at the indexation, integration and retrieval of sound in multimedia documents, and resulted in the creation of DiTaBBu [2], a construction and repurposing framework for DTBs (see figure 1). DiTaBBu integrates a set of media processing modules, in particular, a speech alignment one. This module is responsible for locating the word limits on the speech audio signal and generating the corresponding indexable-media component. That component identifies the entry point in the audio file for each word and, along with the structured text files, constitutes the core of the DTBs. At a later stage (UI-generation), DiTaBBu enables the creation of the actual DTBs, by including navigation, indexing and imaging support, and configuring its presentation and interaction facilities. Several output formats are available, from proprietary ones to DTB standard and SMIL profiles. Details on DiTaBBu can be found in [2].

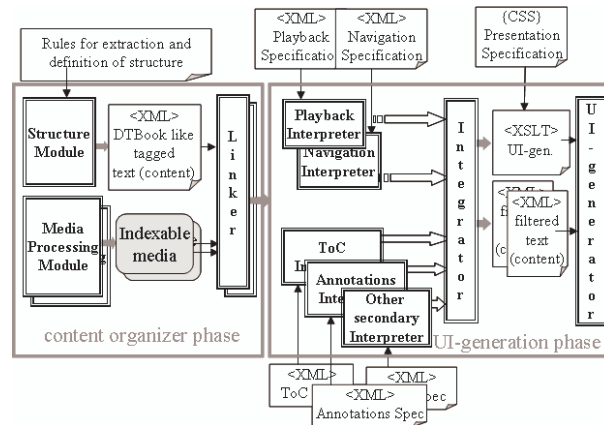


Figure 1: *DiTaBBu - production of Rich DTBs.*

Within IPSOM, the generated book formats aimed, mostly, off-the-shelf players. However, the requirements imposed by the use of specific and diversified technology, by performance issues and by the functionality and demanded features, like the ones reported in this paper, led to the development of dedicated players. RiPlay is one of such players, developed in the context of RiCoBA [7],

an IPSOM continuation project, also involving researchers on speech processing and multimodal interaction, as well as the National Library (owner and publisher of analogue talking books).

## 4 RiPlay

The RiPlay player provides to DTB readers/listeners most of the functionalities proposed in the Playback Device Feature List document [9]. Some of those features are, for example: synchronization between text and audio, highlighting the text unit (word, sentence, ...) being heard at that moment; variable reading speed; visualization and insertion of textual/voice annotations; table of content navigation; text unit jumping, through search and direct choice (e.g. clicking); or even personalizing the user interface by changing font sizes and menu positions.

Special care was taken in usability and accessibility issues, stressing, for instance, the coherence of size in the presentation font of DTB content, indexes and interaction elements; the duration and articulation of audio warnings with narration; and in general the feasibility of using the player functionality without requiring the visual counterpart (audio only). In this perspective, RiPlay provides, not only the ability of work with speech interaction, but it also enables the definition of context-spaces that, in an audio-only version, enables the listener to have the notion of context when navigating through the book (e.g. when a word search occurs, the player reproduces the words from the beginning of the sentence or paragraph, indicating the section or chapter as needed). The player is also highly customizable, enabling, for example, the definition of size of the highlighted text unit and the way annotations and footnotes are notified.

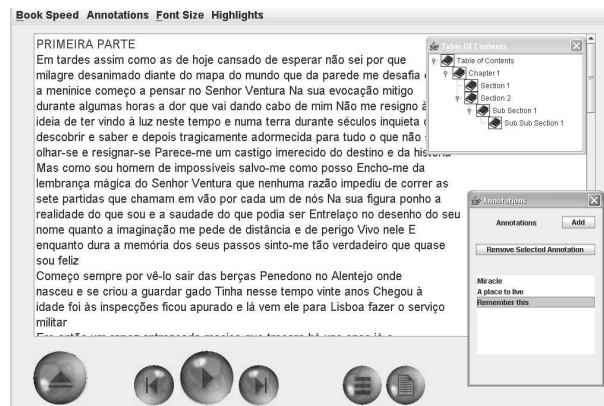


Figure 2: *RiPlay - a Rich DTB player.*

Figure 2 shows a facet of RiPlay with table of contents and the annotations manager opened. Through the latter, new annotations can be added and older ones can be read/listened to, modified or deleted. When added, annotations are associated with the text being reproduced at that moment (narration usually stops, depending on the users' customization). When the player reads that text at a later time, the user is notified that there is an associated annotation. Again, content narration may continue, stop or pause temporarily. Notification can be

done audibly (a sound or through synthesized speech), visually (e.g. highlighting the text) or both. The annotation itself can be opened and reproduced or maintained closed. All these options are configurable.

Although some of these characteristics go, already, beyond the guidelines of DTB and the current implementation of DTB players, two features were introduced that particularly aim at the studying process of text books and articles. These features concern fast overview reading and text re-reading of highlighted parts. Both, in a DTB context, take advantage of the speech reproduction component, thus becoming overview listening and text re-listening. A correlated feature is the creation and reading of text trails that may constitute content perspectives, sub-stories, argumentation paths, etc.

## 4.1 Overview Reading/Listening

Overview reading pertains to the ability to make fast readings of whole texts by picking only specific excerpts according to a selection criterion. Those criteria range from purely syntactic ones, for example, reading the beginning of the first sentence of each section, to more semantic ones, where the reader picks keywords, scans the text for them, and reads the surrounding text.

In view of that, RiPlay provides a choice over several selection criteria that enable different book reproduction. In the simplest one, only the beginning of each paragraph is reproduced. The player plays a few words and then jumps to the beginning of the next paragraph. All parameters can be customized, specifying how many basic units (e.g. words or sentences) will be played in the beginning of each larger unit (e.g. sentence, paragraph, section or chapter). Also, the user can select specific sections (or chapters) to be reproduced, for example, abstracts, introductions or conclusions. Finally, instead of jumping (not playing) the text between the selected excerpts, the user may also opt to play it faster, in order to still have an overall idea of the text.

Other reproduction criteria capitalize on the assumption that the relevant text is around the relevant words. As a basic, yet effective approach [6], RiPlay provides an automatic reproducing mode where the played words are in the neighbourhood of the most used ones: the number of occurrences of significant words is computed; the list of those above a specified frequency is selected; and the sentences where those words occur are reproduced. The frequency threshold can be specified. Still on the same neighbourhood rationale the user has the possibility to define the keywords he wants to search (a relevant words' list). In both cases, the neighbourhood size is customizable and, again, the user may decide that only those words are played or that they are narrated at a smoother pace, whereas the rest of the text is read faster.

## 4.2 Re-reading/listening

Re-reading pertains to the revisiting of a document in order to remember or get a deeper comprehension of specific excerpts considered important on previous readings. The reader usually underlines relevant passages of the text, sometimes, using different colours or marking styles, in order to convey different relevance levels or categories. In a later re-reading, the reader attention is usually focused on those passages.



The RiPlay tool also provides mechanisms for re-reading, again capitalizing on the ability to narrate the content. The user is able to mark excerpts of the text, indicating, currently, three types of excerpts. From a menu the user can choose: *Highlight personal markings*, that marks the text as relevant; *Highlight for fast speed reading*, meaning lesser relevancy; and *Highlight for slow speed reading*, indicating strong importance. Text selection is done as in common text editors/viewers in a visual interaction approach or through voice commands when a speech-based strand is in use. In the latter case the user issues the "begin and end selection" voice commands while listening to the text. As the commands are usually given after the actual words are played, RiPlay marks the text a couple of words before (the number of words is adjustable).

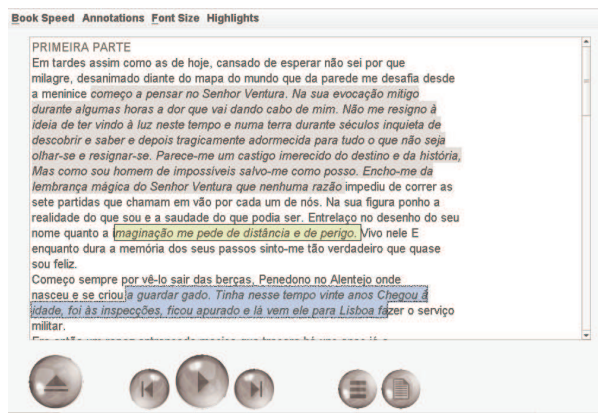


Figure 3: *Highlighting for different speed readings in RiPlay.*

When the text selection is performed, RiPlay visually highlights the text using different colours and different patterns that depend on the relevancy choice that was set (see figure 3). Naturally, it also associates different reading speeds. Colour, patterns and reading speeds are customizable. Once the "highlights" are defined, if the whole document is played, then only the marked excerpts are read. Personal ones are read at the normal pace, whereas the others are read at the specified slow and faster speeds.

Apart from the whole text reading, the user can, for example, read only the text he marked for each of the relevancy choices previously done. Reading options can be changed from the menu shown in figure 4.

### 4.3 Topic re-reading/listening

RiPlay further permits the combination of the above functionalities in order to provide relevant reading/listening over specific topics of document. A simple one, for example, combines the list of relevant words criteria and its neighbourhoods, with the relevancy choices. This way when the document is played, only the neighbourhoods of the words that lay inside selected relevant text are narrated. Moreover they are narrated at the speed defined by the particular selections.

On the other hand, since relevancy selections can be saved and loaded from the player (see figure 4), the user can associate "highlighting" sets to specific

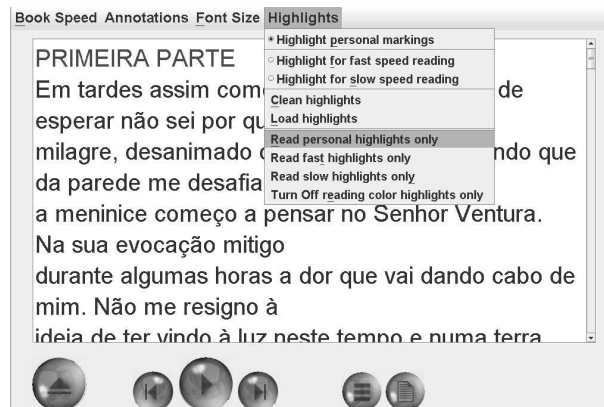


Figure 4: *Types of reading in RiPlay.*

document topics or argumentations.

## 5 Conclusions and Future Work

In this paper we have presented an overview of most recent DTB developments, describing how DTB players have evolved from simple audio readers to complex multimedia systems. We have discussed how newer players can help the user by providing helping mechanisms like the allowing of annotations, markings, interface personalization or book navigation.

However, we have seen how these players still lack flexibility when it comes to the types of readings allowed. The RiPlay player aims to overcome these limitations by providing ways to make overview readings of the text, reading specifically user defined sections and automatically changing book reading speed according to contextual relevance, amongst other new features like sharing annotations and highlights defined by the user, higher interface personalization or even better navigation options.

As future work we are considering some improvements on the mechanism that finds the most important sentences in the book. This mechanism works only as an approximation to the way people traditionally search for important sections and can be greatly improved if we extract the syntactic information from each sentence. Building on these results we can improve our searching algorithm by discarding, for example, the verbs in each sentence or by searching only the names, increasing the accuracy of the results found.

User's annotations can be further developed, with the corresponding enhancements over the control of re-reading/listening. One possibility is allowing the reader to define an annotation hierarchy, and use it to categorize the annotations. By associating playback settings with each category, the re-reading or listening of the book would be further tailored to the user's interests. This could be extended to the use of a metadata language to classify annotations and control future book's reproductions.

To test all the features implemented in the RiPlay player, we are planning a usability evaluation with a special focus in the interface and interaction options available, which will allow us to get a clear idea of the flaws the prototype might

have and help to consolidate the ideas and work ahead.

Research is also being made on the interaction the player allows as we are trying to make this player available to the greatest number of people possible. Future solutions might include for example the connection between the RiPlay player and a braille interface machine, allowing other output options.

## References

- [1] ANSI/NISO. Specifications for the digital talking book. <http://www.niso.org/standards/resources/Z39-86-2002.html>, 2002.
- [2] CARRIÇO, L., DUARTE, C., GUIMARÃES, N., SERRALHEIRO, A., AND TRANCOSO, I. Modular production of flexible digital talking books. In *Proceedings of the 6th International Conference on Enterprise Information Systems, ICEIS 2004* (Porto, Portugal, 2004).
- [3] DUARTE, C., AND CARRIÇO, L. Users and usage driven adaptation of digital talking books. In *Proceedings of the 11th International Conference on Human-Computer Interaction* (Las Vegas, 2005).
- [4] FAIRBAIRN, G. J., AND WINCH, C. *Reading, Writing and Reasoning - A guide for Students*. Open University Press, 1996.
- [5] GAZZANIGA, M. S., IVRY, R. B., AND MANGUN, G. R. *Cognitive Neuroscience - the Biology of the Mind*. W. W. Norton & Company, 1998.
- [6] GERNSBACHER, M. A. *Language Comprehension as Structure Building*. Lawrence Erlbaum, Hillsdale NJ, 1990.
- [7] GUIMARÃES, N., DUARTE, C., CARRIÇO, L., CHAMBEL, T., TEIXEIRA, C., AND SERRALHEIRO, A. RiCoBA - rich content books for all. POSI EIA/61042/2004 - Project Proposal, July 2004.
- [8] LEEUW, E. D., AND LEEUW, M. D. *Read Better, Read Faster*. Penguin Books Lda, 1990.
- [9] NISO. Digital talking book player features list. <http://www.loc.gov/nls/niso/features.htm>, 1999.